

# „Statistical illiteracy“ im Gesundheitswesen am Beispiel des Mammographie-Screenings

Seminararbeit an der Martin-Luther-Universität Halle-Wittenberg im Studiengang Wirtschaftswissenschaften

Tobias Schäffer

Eingereicht im Juni 2018

DOI: <<http://doi.org/10.60019/DBFR5415>>.

## Abstract

Obwohl statistische Informationen allgegenwärtig sind und in verschiedensten Kontexten eingesetzt werden, kann nicht davon ausgegangen werden, dass die Adressaten diese korrekt einschätzen können – es fehlt an „statistical literacy“. Diese umfasst Fähigkeiten der Interpretation, Bewertung und Reflexion statistischer Aussagen, welche im Zusammenspiel mit kognitiven und dispositionellen Komponenten stehen. Frühere Studien zeigten bereits, dass Ärzte nicht fähig waren, Testergebnisse anhand statistischer Informationen korrekt einzuschätzen. Hinzu kommen verzerrte und/oder intransparente Darstellungen in medizinischer Berichterstattung. Am Beispiel des Mammographie-Screenings zeigt diese Arbeit, dass der Zweck von Statistiken, welcher in der objektiven Darstellung empirischer Daten liegt, nur bedingt erfüllt werden kann. Eine informierte Entscheidungsfindung vor allem auf der Seite der PatientInnen ist in diesem Sinne unmöglich. Die konzipierten Lösungsansätze könnten künftig zu größerer „statistical literacy“ im Gesundheitswesen führen.

## Keywords

Gesundheitssystemforschung • Gesundheitswesen • Informationskette • Manipulationsrisiken • Objektivität • Gesundheitsstatistiken • Risikoreduktion • Mortalitätsrate • Mammographie • Statistik

## Gliederung

<b>1</b>	<b>Einleitung .....</b>	<b>9</b>
<b>2</b>	<b>Definition des Forschungsgegenstands .....</b>	<b>10</b>
2.1	Ein einfaches Modell von „statistical literacy“ .....	10
2.2	Das Mammographie-Screening .....	11
<b>3</b>	<b>Vorstellung relevanter Gesundheitsstatistiken.....</b>	<b>13</b>
3.1	Sensitivität und Spezifität eines Screenings .....	13
3.2	Relative und absolute Risikoreduktion.....	13
3.3	Fünfjahresüberlebensrate und Mortalitätsrate .....	14
<b>4</b>	<b>Statistical illiteracy im Gesundheitswesen .....</b>	<b>16</b>
4.1	Das Verständnis von Spezifität und Sensitivität eines Screenings.....	16
4.2	Das Verständnis von relativen und absoluten Risikoreduktionen .....	17
4.3	Das Verständnis von Überlebens- und Mortalitätsraten .....	18
<b>5</b>	<b>Gründe und Lösungsansätze.....</b>	<b>20</b>
5.1	Gründe für „statistical illiteracy“ .....	20
5.1.1	Mangelnde Rechenfähigkeiten.....	20
5.1.2	Unzureichende Ausbildung des medizinischen Nachwuchses .....	20
5.1.3	Verzerzte Berichterstattung in medizinischen Fachzeitschriften .....	21
5.2	Lösungsansätze für „statistical illiteracy“ .....	21
5.2.1	Verbesserung der medizinischen Ausbildung .....	21
5.2.2	Verwendung von transparenten Gesundheitsstatistiken .....	22
5.2.3	Verbindliche Vorgaben für medizinische Fachzeitschriften .....	23
<b>6</b>	<b>Zusammenfassung und Ausblick.....</b>	<b>24</b>
	<b>Anhang .....</b>	<b>30</b>

## 1 Einleitung

Anfang Mai 2018 berichtete die BBC über ein schwerwiegendes Versäumnis des National Health Services (BBC, 2018). Aufgrund eines fehlerhaften Algorithmus erhielten 450.000 Frauen in Großbritannien zwischen 69 und 71 Jahren keine Einladung für eine planmäßige Mammographie. Dieses Versäumnis verkürzte laut National Health Service das Leben von bis zu 270 Frauen. Eine nicht weniger große Debatte löste Angelina Jolie aus, als sie 2015 vorsorglich ihre Brüste amputieren und durch Silikonkissen ersetzen ließ. Grund hierfür war ein Gentest, der ihr eine erhöhte Wahrscheinlichkeit für Brustkrebs bescheinigte. Aufgrund ihres radikalen Beispiels und öffentlichen Bekennnisses erhöhten sich die Anfragen für solche Gentests und Operationen teilweise um das zehnfache (Hengstenberg, 2013). Das erhöhte Bewusstsein war jedoch nicht mit einem besseren Verständnis für die Krankheit oder einem Wissen über die Wahrscheinlichkeit für eine Erkrankung verbunden (Borzekowski et al., 2014).

Nachrichten bzw. Informationen wie diese sind alltäglich. Alle Mitglieder moderner Technologiegesellschaften sind Konsumenten<sup>1</sup> von statistischen Informationen. Durch den Fortschritt in der Informationsverarbeitung wurden rein verbale Äußerungen von Wahrscheinlichkeiten durch numerische Exaktheit abgelöst (Gigerenzer et al., 2005). Dieser Fortschritt führte nicht zu mehr Klarheit, sondern zu mehr Verwirrung. Experten haben nur bedingt gelernt, die exakten Wahrscheinlichkeiten leicht verständlich zu kommunizieren. Statistiken sollten die Freiheit geben, empirische Beweise zu organisieren und objektiv zu evaluieren. Diese Arbeit zeigt am Beispiel des Mammographie-Screenings, dass dieser Zweck aufgrund von „statistical illiteracy“ nur bedingt erfüllt werden kann. Die Unfähigkeit von Ärzten, Patienten, Journalisten und Politikern, mit den entsprechenden Gesundheitsstatistiken umzugehen, macht eine informierte Entscheidungsfindung unmöglich.

Im folgenden Abschnitt wird ein einfaches Modell von „statistical literacy“ erläutert. Anschließend werden das Mammographie-Screening-Programm in Deutschland und hierfür relevante Gesundheitsstatistiken vorgestellt. In Abschnitt 5 wird das Verständnis dieser Gesundheitsstatistiken von Ärzten, Patienten und Politikern überprüft. Darauf aufbauend werden Gründe und Lösungsansätze für die Unfähigkeit mit dem Umgang dieser Statistiken erläutert und die Arbeit mit einer Zusammenfassung der Kernelemente beendet.

---

<sup>1</sup> Diese Arbeit wird zum Wohle der besseren Lesbarkeit nicht gegendert. Wenn nicht explizit angegeben, sind stets Individuen jeglichen Geschlechts gemeint.

## 2 Definition des Forschungsgegenstands

### 2.1 Ein einfaches Modell von „statistical literacy“

Allgemein bezeichnet „literacy“ die Fähigkeit zu lesen und zu schreiben. Die World Literacy Foundation (2015) definiert „complete illiteracy“ als Unfähigkeit, weder schreiben noch lesen zu können. „Functional illiteracy“ beschreibt vorhandene Grundkompetenzen von Lesen, Schreiben und dem Umgang mit Zahlen, welche jedoch nicht ausreichen, um informierte Entscheidungen zu treffen (World Literacy Foundation, 2015).

„Statistical literacy“ ist ein relativ neuer Forschungsbereich und dementsprechend gibt es verschiedene Ansätze zur Definition. Wallman (1993) bezeichnet „statistical literacy“ als Fähigkeit statistische Ergebnisse, die den Alltag durchdringen, zu verstehen und kritisch zu bewerten. Diese sind mit der Fähigkeit gepaart, die Beiträge zu würdigen, die statistisches Denken in öffentlichen und privaten, beruflichen und persönlichen Entscheidungen leisten kann. Gal (2002) definiert „statistical literacy“ als Fähigkeit statistische Informationen, datenbezogene Argumente oder stochastische Phänomene zu interpretieren, kritisch zu bewerten und gegebenenfalls eine Meinung zu äußern.

Watson (1997) entwickelte, ähnlich der Definition der allgemeinen „literacy“, eine dreistufige Hierarchie von „statistical literacy“. Die erste Stufe stellt ein Grundverständnis der probabilistischen und statistischen Terminologie dar. Die zweite Stufe beschreibt ein Verständnis der statistischen Sprache und Konzepte, wenn sie in den Kontext einer breiteren gesellschaftlichen Diskussion eingebettet sind. Die höchste Stufe ist eine fragende Haltung. Diese wird bei der Anwendung von Konzepten genutzt, um Behauptungen zu widersprechen, die ohne angemessene statistische Grundlage aufgestellt wurden (Watson & Callingham, 2003; Watson, 1997).

Darüber hinaus erfordert statistisch gebildetes Verhalten die gemeinsame Aktivierung von kognitiven und dispositionellen Komponenten (Gal, 2005). Die kognitive Komponente umfasst fünf Wissensbasen: Alphabetisierungsfähigkeiten, statistisches Wissen, mathematisches Wissen, Kontext- oder Weltwissen und Wissen über kritische Fragen. Die Dispositionskomponente beinhaltet das Vorhandensein einer kritischen Haltung, d. h. Einstellungen und bestimmte Überzeugungen wie z. B. den Glauben an das Selbst als fähig zum statistischen Denken oder an die Legitimität einer kritischen Sicht auf Informationen, die man von offiziellen Quellen oder von Experten erhält (Gal, 2005). Je nach Ausprägung dieser Elemente ergibt sich, bezogen auf eine bestimmte Fragestellung, eine Stufe der „statistical literacy“. Diese Zusammensetzung wird durch **Abbildung 1** dargestellt.



**Abbildung 1:** Ein einfaches Modell von „statistical literacy“<sup>2</sup>

Das Modell von Gal (2002) wurde um die drei Stufen von „statistical literacy“ nach Watson (1997) erweitert und deckt sich ebenfalls mit der Definition nach Ben-Zvi und Garfield (2004). Die einzelnen

<sup>2</sup> Quelle: Freie Darstellung nach Gal (2002, S. 4) und Watson (1997).

Elemente sollten nicht als alleinstehend bzw. feste Einheiten gesehen werden. Sie stehen vielmehr in einer kontextabhängigen dynamischen Beziehung, wobei die Wissenselemente für das Verständnis sowie die Interpretation notwendig sind (Gal, 2002). Um darüber hinaus Statistiken beurteilen zu können, ist eine kritische Haltung notwendig, die durch das statistische Verständnis und entsprechende Überzeugungen unterstützt wird (Sharma, 2017). Zusammenfassend bedeutet dies, dass die Stufe vom Kontext abhängt, in welchem die Problemstellung auftritt. Ein Individuum kann je nach Fragestellung verschiedene Stufen der „statistical literacy“ aufweisen.

## 2.2 Das Mammographie-Screening

In Deutschland erkranken pro Jahr rund 70.000 Frauen und knapp 700 Männer an Brustkrebs und das Mamma-Karzinom stellt eine der häufigsten Krebserkrankungen von Frauen in Deutschland dar (Robert Koch-Institut, 2016). Jährlich sterben rund 17.000 Patienten an dieser Krankheit. Im internationalen Vergleich zeigt sich, dass Deutschland bezüglich der Prävalenz und Mortalität im Mittelfeld liegt (Robert Koch-Institut, 2016; WHO, 2014).

Eine Mammographie ist eine Röntgenuntersuchung zur Feststellung von Gewebsveränderungen in der Brust. Ein Screening bezeichnet eine Vorsorgeuntersuchung, die ohne medizinische Indikation zur Früherkennung einer Krankheit durchgeführt wird. In Deutschland legt der Gemeinsame Bundesausschuss (G-BA) in seiner Richtlinie zur Krebsfrüherkennung fest, dass Frauen zwischen 50 und 69 Jahren alle 24 Monate einen Anspruch auf ein Mammographie-Screening haben (G-BA, 2017).

Das Mammographie-Screening dient zur „qualitätsgesicherten Früherkennung von Brustkrebs bei Frauen ohne Symptome“, wobei das übergeordnete Ziel die Senkung der brustkrebspezifischen Mortalität ist (Kooperationsgemeinschaft Mammographie, 2017, S. 7). Ein Brustkrebs-Screening-Programm bewahrt je nach Studie zwischen 0,5 und 9 Frauen pro 1.000 behandelter Frauen vor einem Tod durch Brustkrebs (**Tabelle 2**). Im Jahr 2015 nahmen 2.850.458 Frauen in Deutschland am Screening-Programm teil. Die Kosten hierfür betrugen ca. 270 Mio. EUR.<sup>3</sup> 25 der 28 Mitgliedsstaaten der Europäischen Union planen oder führen aktuell ein Screening-Programm für Brustkrebs aus (Basu et al., 2018).

Die potenziellen Schäden einer Mammographie entstehen zunächst in Form einer Strahlenbelastung. Die Wahrscheinlichkeit, durch diese Belastung an Brustkrebs oder einer anderen Krebsart zu erkranken, wird als äußerst gering, jedoch im Vergleich höher als ohne Untersuchung eingestuft (Pauwels, Foray & Bourguignon, 2015). Wahrscheinlicher ist eine falsch-positive oder falsch-negative Diagnose sowie eine Überbehandlung aufgrund des Screenings (Løberg et al., 2015).

Die falsch-positive Testrate im Jahr 2015 lag in Deutschland bei ca. 3,7 % und damit im europäischen Durchschnitt (Kooperationsgemeinschaft Mammographie, 2017). In Deutschland erhielten 123.289 Frauen ein positives Screening-Ergebnis, von welchen 16.785 letztendlich die Diagnose Brustkrebs erhielten. Darüber hinaus werden durch das Screening auch Diagnosen gestellt und Behandlungen eingeleitet, die im Lebensverlauf der Patientinnen ohne eine Vorsorgeuntersuchung nicht notwendig geworden wären. Diese Überbehandlung betrifft je nach Studie zwischen drei und zehn pro 1.000 untersuchten Frauen (Tabelle 2).

Die Behandlung von klinisch nicht-relevanten Erkrankungen erzeugt Kosten ohne dabei einen Nutzen zu stiften (Felder & Mayrhofer, 2017). Zusätzlich werden Patienten physisch und psychisch unnötig belastet.

---

<sup>3</sup> Erläuterung zur Berechnung der Kosten im Anhang.

Es ist somit unklar, ob die imperfekte Spezifität eines Tests weniger schädlich als die imperfekte Sensitivität ist.

In Anbetracht der Tatsache, dass kein medizinischer Test eine perfekte Sensitivität und Spezifität aufweist, sollte die Kausalkette, die durch ein falsch-positives Ergebnis in Gang gesetzt werden kann, nicht banalisiert werden (Aronowitz, 2015).

### 3 Vorstellung relevanter Gesundheitsstatistiken

#### 3.1 Sensitivität und Spezifität eines Screenings

Im Zusammenhang von falsch-positiven und falsch-negativen Testergebnissen eines Screenings spricht man von der Sensitivität (ST) und Spezifität (SP) (Felder & Mayrhofer, 2017). Die ST ist ein statistisches Maß für die Fähigkeit eines Tests, Patienten mit einer Erkrankung korrekt zu identifizieren. Die Berechnung wird anhand der nicht entdeckten klinisch relevanten Erkrankungen gemessen und beträgt beim Mammographie-Screening ca. 90 % (Wegwarth & Gigerenzer, 2011a).

$$ST = \frac{\text{Anzahl der richtig diagnostizierten Patienten}}{\text{Anzahl der tatsächlichen Patienten}}$$

Das bedeutet, dass von zehn an Brustkrebs erkrankten Frauen, neun entdeckt werden. Eine von zehn erkrankten Frauen wird durch das Mammographie-Screening nicht korrekt diagnostiziert.

Die SP wird anhand der falsch-positiven Testergebnisse gemessen, wobei diese als weniger schädlich angesehen werden, da weitere Untersuchungen eine endgültige Abklärung sichern (Felder & Mayrhofer, 2017).

$$SP = \frac{\text{Anzahl der falsch diagnostizierten Patienten}}{\text{Anzahl der tatsächlichen Patienten}}$$

Die falschen Erstdiagnosen führen neben den weiteren Abklärungsbehandlungen jedoch auch zu psychologischem Disstress der Patientinnen.

#### 3.2 Relative und absolute Risikoreduktion

Das Risiko (R) an einer Krankheit zu erkranken, kann wie folgt dargestellt werden (Woloshin et al., 2008, S. 13):

$$R = \frac{\text{Anzahl der diagnostizierten Patienten}}{\text{Anzahl der potentiellen Patienten}}$$

Die Anzahl der Diagnosen beschränkt sich definitionsgemäß auf eine Krankheit. Der Nenner enthält die Personen, die dem Risiko einer Erkrankung potenziell ausgesetzt sind. Das Risiko an einer Krankheit zu sterben, wird als Letalität bezeichnet und ergibt sich, wenn man die Anzahl der Todesfälle in einem bestimmten Zeitraum bzw. -punkt durch die Anzahl der Erkrankten im gleichen Zeitraum bzw. -punkt teilt. Das Ergebnis dieser Berechnung kann sehr unterschiedlich dargestellt werden. Deutlich wird dies anhand nachfolgender Umrechnungstabelle:

1 in _	Dezimal	Prozent	_ von 1.000
1 in 1	1,00	100 %	1.000 von 1.000
1 in 100	0,01	1 %	10 von 1.000
1 in 1.000	0,001	0,1 %	1 von 1.000
1 in 10.000	0,0001	0,01 %	0,1 von 1.000
1 in 100.000	0,00001	0,001 %	0,01 von 1.000

Tabelle 1: Umrechnungstabelle für Risikoangaben<sup>4</sup>

<sup>4</sup> Nach Woloshin et al. (2008, S. 126–127).

Intuitiv wird deutlich, dass Angaben mit vielen Stellen nach dem Komma schwerer verständlich und interpretierbar sind. Ebenso ist die prozentuale Angabe nicht derart leicht verständlich wie die absolute Anzahl der erkrankten Personen in der ersten oder letzten Spalte. Die Nutzung von prozentualen Verhältnissen kann Effekte größer wirken lassen als sie in Wahrheit sind. Deutlich wird dies bei der Angabe von Risikoverminderungen sog. Relative Risikoreduktion (RRR), die durch ein Screening entstehen. Diese Verminderungen können in einem prozentualen bzw. absoluten (ARR) Verhältnis oder als Anzahl der Personen, die behandelt werden müssen, um den Tod einer Person zu vermeiden (NNT=1/ARR), angegeben werden.

Angenommen ein Screening sorgt dafür, dass pro 1.000 behandelter Frauen anstatt vier nun drei Patienten an einer Krankheit sterben. In diesem Fall lässt sich die Risikoreduktion auf drei Arten ausdrücken: Die RRR beträgt 25 %, die ARR beträgt 1 von 1.000 Patienten und die NNT beträgt 1.000. Anhand der **Tabelle 1** lässt sich das absolute Risiko auch als 0,1 % von 1.000 Patienten ausdrücken.

Dieser Unterschied verdeutlicht, dass die Verwendung von relativen Risiken und speziell die Verwendung der RRR ohne die dazugehörige Angabe der ARR die Größe des Effekts verzerrt darstellen kann. Dies ist als „mismatched framing“ bekannt (Gigerenzer et al., 2007). Der Nutzen eines Screenings kann als RRR angegeben werden, da dieser Wert per Definition wesentlich größer ist, wohingegen der Schaden in absoluten Zahlen angegeben werden kann. Diese unterschiedlichen Formate der Darstellung können zu einer Missinterpretation des Nutzens und Schadens eines Screenings führen (Gigerenzer et al., 2007).

### *3.3 Fünfjahresüberlebensrate und Mortalitätsrate*

Die Fünfjahresüberlebensrate (SR) gibt den Prozentsatz der Menschen einer Studie oder Behandlungsgruppe an, die fünf Jahre nach Diagnose oder Beginn der Behandlung noch am Leben sind (WHO, 2018). Die Formel zur Berechnung sieht wie folgt aus:

$$\text{SR} = \frac{\text{Anzahl der Erkrankten, die fünf Jahre nach der Diagnose am Leben sind}}{\text{Anzahl aller mit dieser Krankheit diagnostizierten Patienten}}$$

Die SR gibt dementsprechend die Letalität einer Krankheit innerhalb eines Zeitraumes, in diesem Fall fünf Jahre, an. Je höher der Wert ist, umso mehr Patienten sind am Leben und umso positiver wird die Maßnahme eingeschätzt.

Kritisch ist hierbei, dass sowohl im Zähler als auch im Nenner das Wort Diagnose vorkommt. Wenn versucht wird, mit der SR die Wirksamkeit eines Screenings zu belegen, sind die Werte aus zwei Gründen verzerrt (Welch et al., 2000). Erstens können Diagnosen aufgrund von positiven Testergebnissen eines Screenings oder durch die Abklärung von Symptomen erfolgen. Per Definition erfolgt eine Diagnose durch ein Screening wesentlich früher als eine Diagnose durch Symptome. Dies führt zum sogenannten „lead-time bias“ (Shwartz, 1980). Die Verzerrung entsteht dadurch, dass der Zeitraum, in welchem die Krankheit bekannt ist, verlängert wird, ohne den Zeitpunkt des Todes zu verschieben.

Welch et al. (2000) zeigten, dass bei den 20 häufigsten Krebsarten eine Erhöhung der SR nicht mit einer sinkenden Sterblichkeit korreliert war. Der zweite Grund für eine Verzerrung ist der sog. „overdiagnosis-bias“ (Welch & Black, 2010). Diese Verzerrung entsteht dadurch, dass Screenings Zellveränderungen entdecken, die aufgrund ihrer verlängerten präklinischen Phase oder ihrer fehlenden Neigung zum Fortschritt niemals klinisch signifikant werden würden (Wegwarth & Gigerenzer, 2011b). Sofern diese Patienten in die Analyse eingeschlossen werden, vergrößert sich der Nenner und dementsprechend

der ganze Term. Diese zwei Arten der Verzerrung können zu Missinterpretationen der SR bei der Einführung eines Screenings führen. Die Veränderung der SR erlaubt kein verlässliches Urteil über die Wirksamkeit eines Screenings (Gigerenzer et al., 2007).

Neben der SR kann die Fünfjahresmortalitätsrate (MR) zur Beurteilung der Wirksamkeit einer medizinischen Intervention genutzt werden (Sondik, 1990). Diese drückt aus, wie viele Menschen einer Krankheit innerhalb von fünf Jahren zum Opfer fallen. Hierfür wird die Anzahl der durch eine bestimmte Krankheit verstorbenen Patienten durch die Anzahl aller Personen einer Gruppe geteilt (WHO, 2018):

$$MR = \frac{\text{Krankheitsbedingte Todesfälle innerhalb von 5 Jahren}}{\text{Gesamtpopulation der Patientengruppe}}$$

Die Anzahl der Todesfälle wird nicht durch die diagnostizierten Fälle, sondern durch die Gesamtpopulation einer Patientengruppe geteilt. Durch diese Zusammenfassung der Gesamtpopulation einer Gruppe im Nenner werden beide Verzerrungen der SR ausgeschlossen (Gigerenzer et al., 2007).

## 4 Statistical illiteracy im Gesundheitswesen

### 4.1 Das Verständnis von Spezifität und Sensitivität eines Screenings

Da jedes Screening über eine imperfekte Sensitivität und Spezifität verfügt, ist das Verständnis der Testergebnisse von Ärzten elementar, um Patienten aufzuklären (Wegwarth & Gigerenzer, 2011a). Im Falle eines positiven Testresultates sollten Ärzte die Frage beantworten können, wie hoch die Wahrscheinlichkeit ist, tatsächlich an Brustkrebs erkrankt zu sein. Gigerenzer (2011) präsentierte 160 Gynäkologen am Rande eines Kongresses alle notwendigen Informationen, um diese Frage anhand der bedingten Wahrscheinlichkeit zu errechnen. Die Informationen wurden den Ärzten wie folgt präsentiert (Gigerenzer, 2011, S. 1):

- Die Wahrscheinlichkeit, dass eine Frau an Brustkrebs erkrankt, beträgt 1 %.
- Sofern eine Frau an Brustkrebs erkrankt, beträgt die Wahrscheinlichkeit, dass sie positiv getestet wird, 90 %.
- Eine Frau ohne Brustkrebs hat mit 9 % Wahrscheinlichkeit ein falsch-positives Ergebnis.

Es wurden vier Antworten zur Auswahl gegeben (Gigerenzer 2011, S. 1):

- A. Die Wahrscheinlichkeit, im Falle eines positiven Testes an Brustkrebs erkrankt zu sein, beträgt 81 %.
- B. Von zehn Frauen mit einem positiven Mammographie-Ergebnis haben neun Brustkrebs.
- C. Von zehn Frauen mit positivem Mammographie-Ergebnis hat eine Frau Brustkrebs.
- D. Die Wahrscheinlichkeit für Brustkrebs im Falle eines positiven Screenings liegt bei 1 %.

Die richtige Antwort konnte zum einen durch die angegebenen Informationen als bedingte Wahrscheinlichkeit hergeleitet werden (Gigerenzer, 2011, S. 2):

Wahrscheinlichkeit für Brustkrebs im Falle eines positiven Screenings

$$= \frac{0,01 * 0,9}{0,01 * 0,9 + 0,99 * 0,09} = 0,092$$

Bedingte Wahrscheinlichkeiten sind das gängige Format, in welchem Testergebnisse in medizinischen Fachzeitschriften präsentiert werden (Gigerenzer et al., 2007; Hoffrage et al. 2000; Wegwarth & Gigerenzer, 2011a). Nur 21 % der befragten Ärzte entschieden sich für die richtige Antwort C.

60 % der Ärzte sagten entweder, dass neun von zehn Frauen Brustkrebs haben oder die Wahrscheinlichkeit für Brustkrebs bei 81 % liegt. Die restlichen 19 % nahmen an, dass die Wahrscheinlichkeit bei 1 % liegt. Dies bedeutet, dass 79 % der befragten Ärzte ein positives Mammographie-Ergebnis nicht exakt einschätzen konnten.

Ähnliche Befragungen, bezogen auf ein Screening für Darmkrebs (Hoffrage & Gigerenzer, 1998) und ein Screening für das Down-Syndrom (H. West & Bramwell, 2006), zeigten ebenfalls, dass wenige Ärzte in der Lage waren, positive Testergebnisse richtig zu deuten und entsprechende Wahrscheinlichkeiten herzuleiten. Die Missinterpretation der Statistiken führt zu einer Illusion der Sicherheit der Screening-Maßnahme (Gigerenzer et al., 2007b; Wegwarth et al., 2011). Dies bedeutet, dass der Eindruck entsteht, ein Screening würde zu definitiven Diagnosen führen.

Gemäß den Angaben des Evaluationsberichts der Kooperationsgemeinschaft Mammographie (2017) bestätigt sich das Verhältnis der erwähnten Studien, dass von zehn positiven Screenings neun falsch-positiv sind.

#### *4.2 Das Verständnis von relativen und absoluten Risikoreduktionen*

Fürsprecher des Mammographie-Programms stützen sich auf internationale Studien, die nach Einführung eines flächendeckenden Screenings eine Reduktion der Mortalitätsrate feststellten. Das erste Review zu diesem Thema beschrieb eine RRR von 25 % (Nyström, 1996). Dies bedeutet, dass das Risiko an Brustkrebs zu sterben im Vergleich zu einer Kontrollgruppe, die kein Screening erhielt, um 25 % niedriger ist. Ebenso argumentiert der aktuelle Evaluationsbericht der Kooperationsgemeinschaft Mammographie (2017, S. 7):

In großen randomisierten Studien (RCT) aus den 70er und 80er Jahren wurde für eingeladene Frauen eine Risikoreduktion von 20–25 % nachgewiesen (Andersson et al., 1988; Nystrom et al., 2002; Tabar et al., 2011). Diese Effekte werden auch in aktuellen Meta-Analysen der RCT-Daten bestätigt (Canadian Task Force on Preventive Health Care, 2011; Gotzsche und Jorgensen, 2013). Aktuelle Fallkontrollstudien aus laufenden Screening-Programmen in Europa schätzen die Risikoreduktion für eine Teilnehmerin auf 38–48 % (Paci, 2012; Health Council of the Netherlands, 2014). Vor diesem Hintergrund ist die Umsetzung eines Mammographie-Screening-Programms eine gesundheitspolitisch sinnvolle Maßnahme.

Die Veränderung des Risikos wurde nur als RRR ohne Basisrisiko angegeben. Die untenstehende Tabelle 2 fasst einige der vom Evaluationsbericht zitierten Studien zusammen und vergleicht die relative mit der absoluten Risikoreduktion. Hierbei wird erneut deutlich, dass die Verwendung der RRR den Nutzen im Vergleich zur ARR deutlich größer aussehen lässt, obwohl die grundlegende Information dahinter identisch ist. Eine relative Risikoreduktion von 25 % kann den falschen Eindruck erzeugen, dass von 100 Patientinnen 25 weniger an Brustkrebs sterben. Die entscheidende Frage ist, ob Ärzte diese Unterschiede kennen und verstehen. In einer Erhebung fragte ein Chefarzt der Gynäkologie die Ärzte seiner Abteilung, wie viele Frauen weniger durch die Durchführung eines Screenings sterben, wenn die RRR 25 % (siehe Tabelle 2 (Nyström, 1996)) beträgt (Schüssler, 2005). Die Antworten der Ärzte reichten von einer bis 750 Frauen, die aufgrund eines Screenings nicht sterben müssen.

Diese Frage wurde ebenfalls von Gigerenzer et al. (2007) an 150 Gynäkologen gestellt, wobei die richtige Antwort von vier Möglichkeiten ausgewählt werden musste. 66 % der Ärzte beantworteten die Frage richtig, 16 % sagten 25 Patientinnen, 3 % sprachen sich für 100 Patientinnen aus und 15 % vermerkten, dass 250 Patientinnen weniger sterben würden.

Studie	Mortalität ohne Screening	Mortalität mit Screening	RRR	ARR	In absoluten Zahlen	Überbehandlung
Andersson et al. (1988)	3,3 von 1.000	2,7 von 1.000	20 %	0,06 %	0,6 von 1.000	k. A.
Nyström et al. (1996)	4 von 1.000	3 von 1.000	25 %	0,1 %	1 von 1.000	k. A.
Nyström et al. (2002)	5 von 1.000	4 von 1.000	20 %	0,1 %	1 von 1.000	k. A.
Tabár et al. (2011)	6,6 von 1.000	4,6 von 1.000	30 %	0,2 %	2 von 1.000	3,3 von 1.000
Gotzsche et al. (2013)	30 von 2.000	27–25 von 2.000	10 %–15 %	0,05 %	1 von 2.000	10 von 2.000
Paci et al. (2012)	30 von 1.000	21–23 von 1.000	25 %–31 %	0,7 %–0,9 %	7–9 von 1.000	4 von 1.000

*Tabelle 2: Auswertung der im Evaluationsbericht zitierten Studien*

Wegwarth und Gigerenzer (2011) befragten 20 Gynäkologen zu Nutzen und Schäden einer Mammographie. 17 Gynäkologen beschrieben das Verfahren als sicher, wissenschaftlich gut fundiert und sehr empfehlenswert. Nur sieben Ärzte gaben auf Nachfrage eine Zahl bezüglich der RRR an, die von 20 % bis 50 % reichte, wobei die richtige Antwort bei 20 % lag. Auf Nachfrage zu den eventuellen Schäden gab kein Arzt eine mögliche falsch-positive Diagnose und eine dadurch resultierende Überbehandlung an. Andere Nebenwirkungen wurden als harmlos bezeichnet und nur einer der Ärzte konnte entsprechende Zahlen zu spezifischen Nebenwirkungen angeben. Darüber hinaus überschätzten Ärzte die Wirksamkeit einer Behandlung, sofern die Reduktion des Risikos als RRR angegeben wurde.

Diese Wissenslücke überträgt sich direkt auf die Patienten (Domenighetti et al., 2003). In einer Studie von Gigerenzer, Mata und Frank (2009) waren regelmäßige Arztbesuche nicht mit einem besseren Verständnis der Vor- und Nachteile, sondern mit einer Überschätzung des Nutzens eines Screenings assoziiert.

Perneger und Agoritsas (2011) befragten Krankenhauspatienten und Ärzte zur Einschätzung der Wirksamkeit eines erfundenen neuen Medikaments, wobei dieses das Mortalitätsrisiko senkte und dabei mehr Nebenwirkungen verursachte. Die Teilnehmer wurden randomisiert in Gruppen eingeteilt und jede Gruppe erhielt die Informationen in einem anderen statistischen Format. Aufgrund dieser Informationen sollten die Teilnehmer die Wirksamkeit des Medikaments einschätzen und eine Empfehlung abgeben. Die Entscheidung für eine Empfehlung hing stark vom Format der Darstellung ab (RRR = 93,8 %; ARR = 51,8 %). Zwischen den Empfehlungen der Ärzte und der Patienten gab es hierbei keinen signifikanten Unterschied. Der Effekt des „mismatched framings“ wurde bei Ärzten und Patienten in gleichem Maße sichtbar. Eine vergleichbare Studie, bei der nur Patienten befragt wurden, zeigte den gleichen Einfluss des „framings“ durch die Verwendung von RRR (Malenka et al., 1993).

Auch die Gestalter des Gesundheitswesens sind von „statistical illiteracy“ betroffen. Fahey, Griffiths und Peters (1995) legten den Gesundheitsministerien und -ämtern der Regionen Anglia und Oxford vier Szenarien bezüglich einer Einführung eines Mammographie-Screenings vor. Die Szenarien unterschieden sich nur in der Darstellung der Risikoreduktion, wobei einmal die RRR, einmal die ARR mit den absoluten Zahlen, aus welchen sich die RRR errechnete, und einmal die NNT angegeben wurde. Die Gruppe, die die Information als RRR erhielt, war am ehesten bereit die Maßnahme zu fördern. Dies zeigt, dass Politiker ebenfalls durch ein „mismatched framing“ beeinflussbar sind.

#### *4.3 Das Verständnis von Überlebens- und Mortalitätsraten*

Obwohl die Verzerrungen der SR bei der Einschätzung von Screening-Maßnahmen bekannt und leicht verständlich sind, gibt es immer noch Negativbeispiele in der entsprechenden Literatur. Das aktuelle WHO Positionspapier zum Thema Mammographie (2014) wirbt in der Zusammenfassung damit, dass die SR von Frauen, die in einem fortgeschrittenen Stadium diagnostiziert werden, bei 10–40 % und die SR von Frauen, die innerhalb eines Screenings diagnostiziert werden, bei über 80 % liegt. Dieser Vergleich hat jedoch keinerlei Aussagekraft über die Reduktion der krankheitsspezifischen Mortalität, welche das primäre Ziel eines systematischen Screening-Programms darstellt.

Wegwarth et al. (2012) befragten insgesamt 412 Ärzte, die ambulant oder stationär tätig waren, zum Effekt und Nutzen einer hypothetischen Screening-Maßnahme. Die Ärzte wurden randomisiert in zwei Gruppen eingeteilt, wobei eine Gruppe relevante Informationen (Veränderung der MR von 2 zu 1,6 in 1.000) und eine Gruppe irrelevante Informationen (Veränderung der SR von 69 % zu 99 %) zu der Wirksamkeit des Screenings erhielt. 69 % der Ärzte mit irrelevanten Informationen empfahlen ein Screening, wohingegen nur 23 % der Ärzte der anderen Gruppe den Test befürworteten.

Darüber hinaus wurde in dieser Befragung das allgemeine Wissen über Screening-Statistiken getestet. Hierbei zeigte sich, dass 76 % der Ärzte annahmen, dass beide Statistiken die Wirksamkeit eines Screenings belegen und 47 % aller Ärzte gaben fälschlicherweise an, dass das Auffinden von mehr Krebspatienten im Vergleich zu einer nicht untersuchten Population beweist, dass ein Screening Leben rettet. Eine ähnliche Studie von Wegwarth, Gaissmaier & Gigerenzer (2011) zeigte, dass von 65 befragten Ärzten nur zwei Ärzte den „lead time bias“ und keine Person den „overdiagnosis bias“ erklären konnten. Als die Befragten weitere Informationen über den Nutzen des Screenings in Form von Mortalitätsraten erhielten, schätzten nur 5 % die Maßnahme als wirksam ein. Als die gleiche Information in Form von Überlebensraten verteilt wurde, hielten 51 Ärzte (78 %) das Screening für wirksam. Diese Missinterpretation der SR zeigt, dass ebenfalls ein „mismatched-framing“ durch die Verwendung der SR möglich ist.

## 5 Gründe und Lösungsansätze

### 5.1 Gründe für „statistical illiteracy“

#### 5.1.1 Mangelnde Rechenfähigkeiten

Einer der Hauptgründe für „statistical illiteracy“ ist eine unzureichende Ausprägung der Wissenselemente (Abbildung 1). Besonders fällt hierbei die „basic numeracy“, die grundlegende Fähigkeit mit Zahlen umzugehen, ins Gewicht (Gal, 2002). Ebenso wie bei den Patienten ist die „basic numeracy“ auch bei angehenden Ärzten stark mit der Fähigkeit, Testergebnisse richtig zu interpretieren, verbunden (Schwartz, Woloshin, Black & Welch, 1997). Von 250 befragten Medizinstudenten konnten nur ca. 73 % drei Fragen zu grundlegenden Rechenkenntnissen richtig beantworten (Master et al., 2010). Im Verhältnis zu den Patienten schnitten die Medizinstudenten bei der Befragung etwas besser ab, jedoch gab es einen signifikanten Anteil, der die Zahl 1 nicht zu 0,1 % konvertieren konnte. Die Wahrscheinlichkeit, dass Studenten mit schwachen Rechenkenntnissen Risiken missinterpretierten, war über dreimal so hoch als bei den restlichen Studenten. Dementsprechend zeigte eine weitere Untersuchung, bei der 793 Studenten befragt wurden, dass die kognitive Verzerrung in Form schwacher Rechenfähigkeiten zu weniger kritischen Urteilen führte und die Teilnehmer empfänglicher für unfundierte Informationen machte (West, Toplak & Stanovich, 2008). Zusammengefasst zeigen diese Studien, dass mangelnde grundlegende Rechenfähigkeiten Auswirkungen auf die Einschätzung von Risiken und die Interpretation von Testergebnissen haben.

Eine Studie zu den grundlegenden Kompetenzen von Erwachsenen im internationalen Vergleich zeigte, dass die im Jahr 2012 befragten Deutschen eine leicht überdurchschnittliche mathematische Alltagskompetenz aufweisen (Rammstedt, 2013). Dieser Bericht ist Teil der OECD „Skills Matter“-Umfrage, die im Durchschnitt 22,7 % der Befragten mangelnde Basisrechenkenntnisse bescheinigt (OECD, 2016). Diese mangelnden Kenntnisse reichen nicht aus, um tägliche Aufgaben, wie die Umrechnung von Maßeinheiten oder die Berechnung des Rückgelds an der Kasse, erfolgreich zu lösen.

Bezogen auf das Mammographie-Screening zeigte sich, dass Frauen, die geringe Rechenfertigkeiten aufwiesen, die Informationen zur Risikoreduktion des Mammographie-Screenings nicht richtig einschätzen konnten (Schwartz et al., 1997). Eine weitere Studie zeigte, dass von 895 befragten Frauen nur 19,3 % den Nutzen des Screenings, ausgedrückt in der Anzahl der verhinderten Tode, realistisch einschätzten (Chamot & Perneger, 2001).

#### 5.1.2 Unzureichende Ausbildung des medizinischen Nachwuchses

Bereits 1948 empfahl das British Medical Association Curriculum Committee (1948) Statistik in das erste Studienjahr der medizinischen Ausbildung zu integrieren. Dieser Aufruf wurde jedoch erst 1975 erhört, als Statistik ein obligatorischer Teil der medizinischen Ausbildung in Großbritannien wurde – Zehn Jahre später galt dies auch für Österreich, Ungarn sowie Italien (Altman & Bland, 1991). Für Deutschland sind solche direkten Vorschriften im Ausbildungsteil der Approbationsordnung für Ärzte nur indirekt zu finden. Hier heißt es, dass „die für das ärztliche Handeln erforderlichen Kenntnisse...“ vermittelt werden sollen (§1 Abs. 1 S. 4 ÄApprO).

Mittlerweile ist davon auszugehen, dass auch in Deutschland Medizinstudenten in quantitativen Methoden unterrichtet werden. Obligatorische Kurse in Biostatistik vermitteln jedoch keine ausreichenden Kenntnisse der Risikokommunikation (Wegwarth & Gigerenzer, 2011a). Darüber hinaus ergab eine Befragung von 135 medizinischen Fakultäten in den USA, dass im besten Fall ca. 3 % der Fragen im Rahmen der Abschlussprüfung aus dem Bereich klinische Epidemiologie stammen (Wegwarth &

Gigerenzer, 2011a). Risikokommunikation war in keiner der Fakultäten ein Bestandteil der Ausbildung. Eine Weiterbildung im Umfang von 75 Minuten hat bereits zu einem besseren Verständnis der hier vorgestellten Gesundheitsstatistiken geführt (Gigerenzer & Kober, 2013).

Die eingangs erwähnten 160 Gynäkologen, welche die Sensitivität einer Mammographie einschätzen sollten, wurden nach einer ersten Befragung 75 Minuten lang in Risikokommunikation geschult. Durch diese Schulung im Umgang mit natürlichen Häufigkeiten konnte die Anzahl der richtigen Antworten von 21 % auf 87 % gesteigert werden. Dies zeigt, dass das Problem weniger an der kognitiven Beschränkung als an mangelnder Ausbildung und dem Format der Information liegt.

### *5.1.3 Verzerrte Berichterstattung in medizinischen Fachzeitschriften*

In der medizinischen Berichterstattung ist die Nutzung von intransparenten Gesundheitsstatistiken weit verbreitet (Wegwarth & Gigerenzer, 2011a). Wie gezeigt, führt die Verwendung von verschiedenen Formaten für Nutzen (RRR) und Schäden (ARR) einer Maßnahme zur Überschätzung des Nutzens und zur Unterschätzung der Schäden eines Screenings. Eine Analyse von 359 Artikeln, die über randomisierte Kontrollstudien berichteten und in fünf „high impact“ Fachzeitzeitschriften für Medizin<sup>5</sup> erschienen, zeigte ein deutliches Übergewicht der Nutzung von RRR (J Nuovo, Melnikow & Chang, 2002). Nur 25 Studien gaben die ARR an, um die Wirksamkeit der Intervention darzustellen. Die restlichen 93 % gaben die Ergebnisse in einem intransparenten Format an.

Dieses Ergebnis wurde von einer zweiten Analyse bestätigt (Schwartz et al., 2006). Dieses Mal wurden 222 Artikel untersucht, von welchen 68 % keine Angaben zu den absoluten Risiken machten. Eine weitere Analyse von randomisierten Studien der gleichen fünf Journale zeigte, dass es eine große Variation bei der Beschreibung von Schäden der Interventionen gab (Jim Nuovo & Sather, 2007). Von 521 Studien über randomisierte Kontrollstudien berichteten 328 im Abstract, 380 Studien im Methodenteil und 464 Studien im Ergebnisteil von Nebenwirkungen. Die Autoren beschrieben diese Varianz als inkonsistent und verbessерungsbedürftig.

Eine ähnliche Untersuchung für das Fachgebiet Psychologie ergab ebenfalls, dass die Schäden nicht vollständig und systematisch in den randomisierten Studien berichtet werden (Jonsson et al., 2014). Von 132 enthielten 28 Studien (21 %) Informationen, die auf eine Überwachung von Schäden auf Patientenebene hindeuteten. Nur vier dieser 28 Studien lieferten eine vollständige Beschreibung der Nebenwirkungen und erläuterten die Methoden zur Erhebung dieser Daten. Dies zeigt, dass die Verwendung von RRR durch Ärzte, Patienten, Journalisten und Patientenbroschüren dadurch kommt, dass diese Statistik der Standard in der medizinischen Berichterstattung ist (Wegwarth & Gigerenzer, 2011a).

## *5.2 Lösungsansätze für „statistical illiteracy“*

### *5.2.1 Verbesserung der medizinischen Ausbildung*

Die dargestellten Gründe für den fehlerhaften Umgang mit Statistiken lassen den Schluss zu, dass „Statistical illiteracy“ vorrangig kein Problem kognitiver Beschränktheit ist. Die Unwissenheit bezüglich der Unterschiede zwischen relativer und absoluter Risikoreduktion ermöglicht das „mismatched framing“. Daher sollten die Lehrpläne für das Medizinstudium einen größeren Wert auf die statistische Ausbildung von Ärzten legen. Jede medizinische Fakultät sollte ihren Studenten beibringen, wie man Evidenz

<sup>5</sup> Eingeschlossen waren die Fachzeitschriften Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, The Lancet, und das New England Journal of Medicine.

im Allgemeinen und Gesundheitsstatistiken im Besonderen versteht. Darüber hinaus sollte die statistische Kompetenz in der medizinischen Weiterbildung (CME) bewertet werden (Wegwarth & Gigerenzer, 2011a). Medizinstudenten müssen lernen, dass sie AAR, Mortalitätsraten und Zahlen statt rein verbaler Beschreibungen im Rahmen des Screenings verwenden sollten, um Risiken ausreichend bewerten und kommunizieren zu können.

Angehende Mediziner müssen ebenfalls lernen, dass sie selbst und Patienten die Sensitivität und Spezifität eines Screenings anhand von natürlichen Häufigkeiten besser verstehen als anhand von bedingten Wahrscheinlichkeiten (Gigerenzer et al., 2007).

Statistische Kompetenz erfordert auch ein Umdenken in der Lehre der Statistik an medizinischen Fakultäten (Wegwarth & Gigerenzer, 2011a). Anstatt die Statistik zu einem mathematischen Verfahren an sich zu machen, sollte sie als nützliches Werkzeug zur Lösung klinisch relevanter Probleme anhand konkreter Beispiele vermittelt werden.

Auf der anderen Seite sollten Bürger zu mündigen Patienten geformt werden, die keine unnötigen Tests oder Behandlungen anfordern. Daher sollten Grund- und weiterführende Schulen anfangen, das statistische Denken zu lehren (Wegwarth & Gigerenzer, 2011a). Hierzu könnte der bisherige Mathe- matikunterricht um eine Einheit zum Thema Mathematik der Unsicherheit ergänzt werden.

### *5.2.2 Verwendung von transparenten Gesundheitsstatistiken*

Die Unzulänglichkeiten der Berichterstattung über randomisierte Kontrollstudien in medizinischen Fachzeitschriften führten bereits 1996 zur Veröffentlichung der sogenannten CONSORT-Richtlinien (Begg, 1996). Die Forderung nach der Angabe von transparenten Statistiken wurde erst 2010 in eine Neuauflage dieser Richtlinien integriert (Moher et al., 2010). Sinnvolle Alternativen zu den intransparenten Gesundheitsstatistiken sind vorhanden und müssen stärker bei der Berichterstattung und Pati- enteninformation berücksichtigt werden.

Die Vermeidung von bedingten Wahrscheinlichkeiten und Nutzung von natürlichen Häufigkeiten erleichtert die Deutung der Sensitivität und Spezifität eines Screenings (Wegwarth & Gigerenzer, 2011a). Wie hoch die Wahrscheinlichkeit ist, im Falle eines positiven Mammographie-Programms an Brustkrebs erkrankt zu sein, lässt sich wie folgt ausdrücken (Gigerenzer, 2011, S. 2):

$$\text{Wahrscheinlichkeit für Brustkrebs im Falle eines positiven Screenings} = \frac{9}{9 + 89} = 0,092$$

Die hierfür notwendigen Angaben sind, dass von 1000 Frauen zehn an Brustkrebs erkranken. Von die- sen zehn Frauen werden neun durch eine richtig-positive Mammographie diagnostiziert und von den restlichen 990 Frauen erhalten 89 ein falsch-positives Ergebnis. Dies ergibt, dass insgesamt 98 Frauen positiv getestet werden, von welchen neun tatsächlich an Brustkrebs erkrankt sind. Dies entspricht ebenfalls einer Wahrscheinlichkeit von 9,2 %.

Woloshin und Schwartz (2011) fanden anhand einer randomisierten Befragung heraus, dass Ärzte und Patienten natürliche Häufigkeiten besser als Wahrscheinlichkeiten verstanden. Ebenso wurden positive Vorhersagewerte anhand dieses Formats besser von Laien und Medizinstudenten verstanden (Gigerenzer & Hoffrage, 1995). Zhu und Gigerenzer (2006) demonstrierten, dass die Leistung von Sechstklässlern mit natürlichen Häufigkeiten der Fähigkeit von Erwachsenen mit Wahrscheinlichkeiten entsprach. Zusätzlich sollten relative Risikoreduktionen nicht alleinig genutzt, sondern um die Angabe der entsprechenden absoluten Risikoreduktionen ergänzt werden, damit die Ausmaße eines „mis- matched framing“ begrenzt werden können.

Da die „statistical literacy“ bei jedem Individuum kontextabhängig ist, müssen Ärzte versuchen, den Bedürfnissen der Patienten gerecht zu werden. Die richtige Darstellungsform für Patienten zu wählen, ist eine Schlüsselqualifikation von Ärzten bei der Risikokommunikation. Hierbei liegt die größte Hürde vorrangig in der mangelhaften Schulung der Ärzte (Wegwarth & Gigerenzer, 2011a).

### *5.2.3 Verbindliche Vorgaben für medizinische Fachzeitschriften*

Im Jahr 1996 erschien die erste Version der „Consolidated Standards of Reporting Trials“ (CONSORT) (Begg, 1996). Die Autoren entwickelten Mindestanforderungen für die Berichterstattung von randomisierten Kontrollstudien. Im Jahr 2010 erfolgte die letzte Aktualisierung dieser Richtlinien (Moher et al., 2010). Die Umsetzung wird von zahlreichen medizinischen „high impact“ Fachzeitschriften gefördert. Dies sind die gleichen Fachzeitschriften, die durch verzerrte Veröffentlichung negativ auffielen und so das „mismatched framing“ ermöglichten. Studien zur Wirksamkeit der Richtlinien zeigen dementsprechend ein unzureichendes Niveau der Umsetzung (Prady et al., 2008; Shamseer et al., 2016; Turner et al., 2012).

Die Umsetzung der transparenten Berichterstattung sollte daher von den Herausgebern und Referees der Fachzeitschriften mit größerem Druck vorangetrieben werden. Die Verwendung von intransparenten Gesundheitsstatistiken sollte nicht nur im Nachhinein bemängelt werden, sondern zu einer Ablehnung der Artikel führen (Wegwarth & Gigerenzer, 2011a). Die Richtlinien sollten daher in verbindliche Vorgaben umgewandelt werden. Artikel, die diese Vorgaben nicht einhalten, sollten nicht veröffentlicht werden.

## 6 Zusammenfassung und Ausblick

Zweck dieser Arbeit war es nicht, die Mammographie als solche in Frage zu stellen. Vielmehr ging es darum zu zeigen, dass Frauen aufgrund der kollektiven „statistical illiteracy“ keine Möglichkeit haben, den Nutzen und die potentiellen Schäden möglichst objektiv abzuwägen. Das eingangs erwähnte Beispiel der BBC, bestätigt, dass Journalisten Gesundheitsstatistiken entweder selbst nicht verstehen oder in intransparenter Weise wiedergeben. Wie gezeigt, erhöht ein Mammographie-Screening nicht die Lebenserwartung von Frauen, sondern senkt die brustkrebspezifische Mortalität. Zusätzlich blieb in dem Artikel unerwähnt, wie viele Frauen vor einer Überdiagnose und -behandlung bewahrt wurden. Diese Missinterpretationen entstehen durch die kollektive Missinterpretation von Statistiken.

Diese Arbeit hat jedoch gezeigt, dass das Problem weniger bei stabilen kognitiven Defiziten liegt, als in der Art und Weise wie Informationen den Ärzten und Patienten präsentiert werden. Vorrangig sind unzureichende situationsbezogene Wissenselemente und die standardmäßige Verwendung intransparenter Statistiken in medizinischen Fachzeitschriften der Auslöser. Die Lösungsansätze hierfür sind relativ leicht zu realisieren und werden beispielsweise in neuen Informationsbroschüren und den CONSORT-Richtlinien teilweise umgesetzt.

Doch selbst wenn die statistische Kompetenz der allgemeinen Bevölkerung zunimmt und sich die Berichterstattung in Zeitschriften und Broschüren ändert, werden Ärzte und Leistungserbringer die wichtigsten Quellen für Gesundheitsinformationen für einen Patienten bleiben. Auch in Zukunft werden Ärzte die Verantwortung tragen, Patienten dabei zu helfen den potenziellen Nutzen und Schaden der medizinischen Prävention zu verstehen. Es liegt nicht nur in der Verantwortung der medizinischen Ausbildungseinrichtungen, die zweifellos die Qualität der medizinischen Statistik verbessern müssen, dass zukünftige Patienten die Beratung erhalten, die sie erhalten sollen. Es bleibt auch die Verantwortung aller Medizinstudenten und aller zugelassenen Ärzte, neugierig auf statistische Fragen zu bleiben, die sie noch nicht verstanden haben, und auf eine angemessene statistische Ausbildung ihrer medizinischen Fakultäten und Organisationen zum Wohle ihrer Patienten zu bestehen.

Zukünftige Untersuchungen sollten „statistical illiteracy“ im Rahmen von „health literacy“ betrachten, weitere Darstellungsoptionen für Risiken und möglichst verständliche Risikokommunikationsmöglichkeiten evaluieren. Ebenso sollte auf das Problem der Polarisation und den „publication-bias“ bei der Veröffentlichung von Studienergebnissen und die Umsetzung der CONSORT-Richtlinien eingegangen werden.

## Literaturverzeichnis

- Altman, D. G. & Bland, J. M. (1991). Improving Doctors' Understanding of Statistics. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154(2), 223. DOI: 10.2307/2983040.
- Andersson, I., Aspegren, K., Janzon, L., Landberg, T., Lindholm, K., Linell, F., Ljungberg, O., Ranstam, J. & Sigfusson, B. (1988). Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ*, 297(6654), 943–948. DOI: 10.1136/bmj.297.6654.943.
- Aronowitz, R. A. (2015). *Risky medicine: our quest to cure fear and uncertainty*. The University of Chicago Press.
- Basu, P., Ponti, A., Anttila, A., Ronco, G., Senore, C., Vale, D. B., Segnan, N., Tomatis, M., Soerjomataram, I., Primic Žakelj, M., Dillner, J., Elfström, K. M., Lönnberg, S., Sankaranarayanan, R. (2018). Status of implementation and organization of cancer screening in The European Union Member States—Summary results from the second European screening report. *International Journal of Cancer*, 142(1), 44–56. DOI: 10.1002/ijc.31043.
- BBC (Hrsg.). (2018). *Breast screening error ‘shortened up to 270 lives’ – Hunt*. Online: <https://www.bbc.com/news/health-43973652>.
- Begg, C. (1996). Improving the Quality of Reporting of Randomized Controlled Trials. *JAMA*, 276(8), 637. DOI: 10.1001/jama.1996.03540080059030.
- Ben-Zvi, D. & Garfield, J. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. In: D. Ben-Zvi, D. & J. Garfield (Hrsg.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Dordrecht: Springer Science+Business Media, 3–15. DOI: 10.1007/1-4020-2278-6\_1.
- Borzekowski, D. L. G., Guan, Y., Smith, K. C., Erby, L. H. & Roter, D. L. (2014). The Angelina effect: Immediate reach, grasp, and impact of going public. *Genetics in Medicine*, 16(7), 516–521. DOI: 10.1038/gim.2013.181.
- Chamot, E. & Perneger, T. V. (2001). Misconceptions about efficacy of mammography screening: A public health dilemma. *Journal of Epidemiology and Community Health*, 55(11), 799–803. DOI: 10.1136/jech.55.11.799.
- Domenighetti, G., D'Avanzano, B., Egger, M., Berrino, F., Perneger, T., Mosconi, P. & Zwahlen, M. (2003). Women's perception of the benefits of mammography screening: Population-based survey in four countries. *International Journal of Epidemiology*, 32(5), 816–821. DOI: 10.1093/ije/dyg257.
- Fahey, T., Griffiths, S. & Peters, T. J. (1995). Evidence based purchasing: Understanding results of clinical trials and systematic reviews. *BMJ Clinical Research*, 311(7012), 1056–1059. DOI: 10.1136/bmj.311.7012.1056.
- Felder, S. & Mayrhofer, T. (2017). *Medical Decision Making. A Health Economic Primer*. 2. Aufl. Berlin (u. a.): Springer-Verlag. DOI: 10.1007/978-3-662-53432-8.
- Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70(1), 1–25. DOI: 10.1111/j.1751-5823.2002.tb00336.x.

- Gal, I. (2005). Towards Probability Literacy for all Citizens: Building Blocks and Instructional Dilemmas. In: G. A. Jones (Hrsg.). *Exploring Probability in School: Challenges for Teaching and Learning*, 39–63. DOI: 10.1007/0-387-24530-8\_3.
- G-BA (Hrsg.). (2018). *Richtlinie des Gemeinsamen Bundesausschusses über die Früherkennung von Krebserkrankungen, Stand Juli 2018 (Krebsfrüherkennungs-Richtlinie/KFE-RL)*. Berlin: Gemeinsamer Bundesausschuss. Online: [https://www.g-ba.de/downloads/62-492-1811/KFE-RL\\_2018-07-19\\_iK-2019-04-18.pdf](https://www.g-ba.de/downloads/62-492-1811/KFE-RL_2018-07-19_iK-2019-04-18.pdf).
- Gigerenzer, G. & Kober, H. (2013). *Risiko: Wie man die richtigen Entscheidungen trifft*. München: C. Bertelsmann.
- Gigerenzer, G. (2011). What are natural frequencies? *BMJ*, 343. DOI: 10.1136/bmj.d6386.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M. & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, 8(2), 53–96. DOI: 10.1111/j.1539-6053.2008.00033.x.
- Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B. & Katsikopoulos, K. V. (2005). “A 30% chance of rain tomorrow”: How does the public understand probabilistic weather forecasts? *Risk Analysis*, 25(3), 623–629. DOI: 10.1111/j.1539-6924.2005.00608.x.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684–704. DOI: 10.1037/0033-295X.102.4.684.
- Gigerenzer, G., Mata, J. & Frank, R. (2009). Public Knowledge of Benefits of Breast and Prostate Cancer Screening in Europe. *Journal of the National Cancer Institute*, 101(17), 1216–1220. DOI: 10.1093/jnci/djp237.
- Gøtzsche, P. C. & Jørgensen, K. J. (2013). Screening for breast cancer with mammography (Review). *Cochrane Database of Systematic Reviews*, (6). DOI: 10.1002/14651858.CD001877.pub5.
- Hoffrage, U. & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine: Journal of the Association of American Medical Colleges*, 75(5), 538–540. DOI: 10.1097/00001888-199805000-00024.
- Hoffrage, U., Lindsey, S., Hertwig, R. & Gigerenzer, G. (2000). Communicating statistical information. *Science*, 290(5500), 2261–2262. DOI: 10.1126/science.290.5500.2261.
- Jonsson, U., Alaie, I., Parling, T. & Arnberg, F. K. (2014). Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. *Contemporary Clinical Trials*, 38(1), 1–8. DOI: 10.1016/j.cct.2014.02.005.
- Kooperationsgemeinschaft Mammographie (Hrsg.). (2017). *Jahresbericht Evaluation 2015. Deutsches Mammographie-Screening Programm*. Online: [https://newsroom.mammo-programm.de/download/fachpublikation/KOOPMAMMO\\_Jahresbericht\\_EVAL2015\\_20170926\\_web.pdf](https://newsroom.mammo-programm.de/download/fachpublikation/KOOPMAMMO_Jahresbericht_EVAL2015_20170926_web.pdf).
- KV Schleswig Holstein (Hrsg.). (o. J.). *Das Mammographie-Screening-Programm. Informationen für Ärzte*. Online: <https://www.kvsh.de/index.php?StoryID=420>.
- Løberg, M., Lousdal, M. L., Brethauer, M. & Kalager, M. (2015). Benefits and harms of mammography screening. *Breast Cancer Research: BCR*, 17(1), 63. DOI: 10.1186/s13058-015-0525-z.

- Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W. & Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of General Internal Medicine*, 8(10), 534–538. DOI: 10.1007/BF02599636.
- Master, V., Johnson, T., Abbasi, A., Schoenberg, E., Kellum, R., Speake, L., Spiker, C., Foust, A., Kreps, A. & Ritenour, C. (2010). 12 NUMERACY AMONG MEDICAL STUDENTS: ARE WE PREPARING STUDENTS FOR EVIDENCE-BASED MEDICINE? *The Journal of Urology*, 183(4), e5. DOI: 10.1016/j.juro.2010.02.055.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M. & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 63(8), e1-37. DOI: 10.1136/bmj.c869.
- Nuovo, J., Melnikow, J. & Chang, D. (2002). Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *Jama*, 287(21), 2813–2814. DOI: 10.1001/jama.287.21.2813.
- Nuovo, J. & Sather, C. (2007). Reporting adverse events in randomized controlled trials. *Pharmacoepidemiology and Drug Safety*, 16(3), 349–351. DOI: 10.1002/pds.1310.
- Nyström, L. (1996). An overview of the Swedish randomised mammography trials: Total mortality pattern and the representivity of the study cohorts. *Journal of Medical Screening*, 3(2), 85–87. DOI: 10.1177/096914139600300208.
- Nyström, L., Andersson, I., Bjurstrom, N., Frisell, J., Nordenskjöld, B. & Rutqvist, L. E. (2002). Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. *Lancet*, 359(9310), 909–919. DOI: 10.1016/S0140-6736(02)08020-0.
- OECD (Hrsg.). (2016). *Skills Matter: Further Results from the Survey of Adult Skills*. (OECD Skills Studies). Paris: OECD Publishing. DOI: 10.1787/9789264258051-en.
- Paci, E., Broeders, M., Hofvind, S. & Duffy, S. W. (2012). Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *Journal of Medical Screening*, 2012;19 Suppl 1, 5–13. DOI: 10.1258/jms.2012.012077.
- Perneger, T. V. & Agoritsas, T. (2011). Doctors and patients' susceptibility to framing bias: A randomized trial. *Journal of General Internal Medicine*, 26(12), 1411–1417. DOI: 10.1007/s11606-011-1810-x.
- Prady, S. L., Richmond, S. J., Morton, V. M. & MacPherson, H. (2008). A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE*, 13;3(2), e1577. DOI: 10.1371/journal.pone.0001577.
- Rammstedt, B. (Hrsg.) (2013). *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich. Ergebnisse von PIAC 2012*. Münster: Waxmann.
- Robert Koch-Institut (Hrsg.). (2016). *Bericht zum Krebsgeschehen in Deutschland 2016*. DOI: 10.17886/rkipubl-2016-014.
- Schüssler, B. (2005). Im Dialog: Ist Risiko überhaupt kommunizierbar, Herr Prof. Gigerenzer? *Frauenheilkunde Aktuell*, 14, 25–31.

- Schwartz, L. M., Woloshin, S., Black, W. C. & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127(11), 966–72. DOI: 10.7326/0003-4819-127-11-199712010-00003.
- Schwartz, L. M., Woloshin, S., Dvorin, E. L. & Welch, H. G. (2006). Ratio measures in leading medical journals: Structured review of accessibility of underlying absolute risks. *BMJ*, 333(7581), 1248. DOI: 10.1136/bmj.38985.564317.7C.
- Shamseer, L., Hopewell, S., Altman, D. G., Moher, D. & Schulz, K. F. (2016). Update on the endorsement of CONSORT by high impact factor journals: A survey of journal “Instructions to Authors” in 2014. *Trials* 17(301). DOI: 10.1186/s13063-016-1408-z.
- Sharma, S. (2017). Definitions and models of statistical literacy: a literature review. *Open Review of Educational Research*, 4(1), 118–133. DOI: 10.1080/23265507.2017.1354313.
- Shwartz, M. (1980). Estimates of lead time and length bias in a breast cancer screening program. *Cancer*, 46(4), 844–851. DOI: 10.1002/1097-0142(19800815)46:4<844::AID-CNCR2820460434>3.0.CO;2-5.
- Sondik, E. J. (1990). Measurement of progress against cancer. Extramural Committee to Assess Measures of Progress Against Cancer. *Journal of the National Cancer Institute*, 82(10), 825–835.
- Hengstenberg, M. (2013). Genetisches Brustkrebsrisiko – Der Jolie-Effekt. *Spiegel Online*. Online: <https://www.spiegel.de/gesundheit/diagnose/brust-op-bei-angelina-jolie-viele-frauen-sind-besorgt-a-904556.html>.
- Tabár, L., Vitak, B., Chen, T. H., Yen, A. M., Cohen, A., Tot, T., Chiu, S. Y., Chen, S. L., Fann, J. C., Rosell, J., Fohlin, H., Smith, R. A. & Duffy, S. W. (2011). Swedish Two-County Trial: Impact of Mammographic Screening on Breast Cancer Mortality during 3 Decades. *Radiology*, 260(3), 658–663. DOI: 10.1148/radiol.11110469.
- Turner, L., Shamseer, L., Altman, D. G., Schulz, K. F. & Moher, D. (2012). Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Systematic Reviews*, 1(60). DOI: 10.1186/2046-4053-1-60.
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1–8. DOI: 10.2307/2290686.
- Watson, J. & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46.
- Watson, J. M. (1997). Assessing Statistical Thinking Using the Media. In: I. Gal, & J.B. Garfield (Hrsg.). *The Assessment Challenge in Statistics Education*. IOS Press, 107–121.
- Wegwarth, O., Gaissmaier, W. & Gigerenzer, G. (2011). Deceiving numbers: Survival rates and their impact on doctors’ risk communication. *Medical Decision Making*, 31(3), 386–394. DOI: 10.1177/0272989X10391469.
- Wegwarth, O. & Gigerenzer, G. (2011a). Statistical Illiteracy in Doctors. In: G. Gigerenzer & M. Gray (Hrsg.). *Better Doctors, Better Patients, Better Decisions. Envisioning Health Care 2020*. MIT Press, 137–152.

- Wegwarth, O. & Gigerenzer, G. (2011b). “There is nothing to worry about”: Gynecologists’ counseling on mammography. *Patient Education and Counseling*, 84(2), 251–256. DOI: 10.1016/j.pec.2010.07.025.
- Wegwarth, O., Schwartz, L. M., Woloshin, S., Gaissmaier, W. & Gigerenzer, G. (2012). Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Annals of Internal Medicine*, 156(5), 340–349. DOI: 10.1059/0003-4819-156-5-201203060-00005.
- Welch, H. G. & Black, W. C. (2010). Overdiagnosis in Cancer. *JNCI Journal of the National Cancer Institute*, 102(9), 605–613. DOI: 10.1093/jnci/djq099.
- Welch, H. G., Schwartz, L. M. & Woloshin, S. (2000). Are increasing 5-year survival rates evidence of success against cancer? *JAMA*, 283(22), 2975–2578. DOI: 10.1001/jama.283.22.2975.
- West, H. & Bramwell, R. (2006). Do maternal screening tests provide psychologically meaningful results? Cognitive psychology in an applied setting. *Journal of Reproductive and Infant Psychology*, 24(1), 61–69. DOI: 10.1080/02646830500475278.
- West, R. F., Toplak, M. E. & Stanovich, K. E. (2008). Heuristics and Biases as Measures of Critical Thinking: Associations with Cognitive Ability and Thinking Dispositions. *Journal of Educational Psychology*, 100(4), 930–941. DOI: 10.1037/a0012842.
- Woloshin, S. & Schwartz, L. M. (2011). Communicating data about the benefits and harms of treatment. *Annals of Internal Medicine*, 155(2), 87–96. DOI: 10.7326/0003-4819-155-2-201107190-00004.
- Woloshin, S., Schwartz, L. M. & Welch, H. G. (2008). *Know Your Chances: Understanding Health Statistics*. University of California Press.
- WHO (Hrsg.). (2014). *WHO Position Paper on Mammography Screening*. Genf: World Health Organization. Online: [https://apps.who.int/iris/bitstream/handle/10665/137339/9789241507936\\_eng.pdf;jsessionid=61BB7A03DABAF5C8559116D5A20119CC?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/137339/9789241507936_eng.pdf;jsessionid=61BB7A03DABAF5C8559116D5A20119CC?sequence=1).
- Zhu, L. & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, 98(3), 287–308. DOI: 10.1016/j.cognition.2004.12.003.

## Anhang

Auf Grundlage der verfügbaren Zahlen des Evaluationsberichtes der Kooperationsgemeinschaft Mammographie (2017) wurden mithilfe der EBM-Kennziffern die ungefähren Kosten des Mammographie-Screenings in Deutschland im Jahr 2015 geschätzt.

Im ersten Schritt wurden die betreffenden Geldwerte der EBM-Kennziffern den Ablaufschritten des Mammographie-Screening-Programms zugeordnet (KV Schleswig Holstein, o. J.). Anschließend wurden die Zahlen des Evaluationsberichtes genutzt, um so die Gesamtsumme zu ermitteln (siehe **Abbildung 2**). Im Evaluationsbericht waren folgende Angaben bezüglich der Mengen vorhanden:

Eingeladene Teilnehmerinnen:	5.528.937
Durchgeführte Mammographien:	2.850.458
Wiedereinbestellung (positive Mammographien):	123.289
Diagnose Brustkrebs:	16.785
Diagnose durch Stanzbiopsie:	15.625

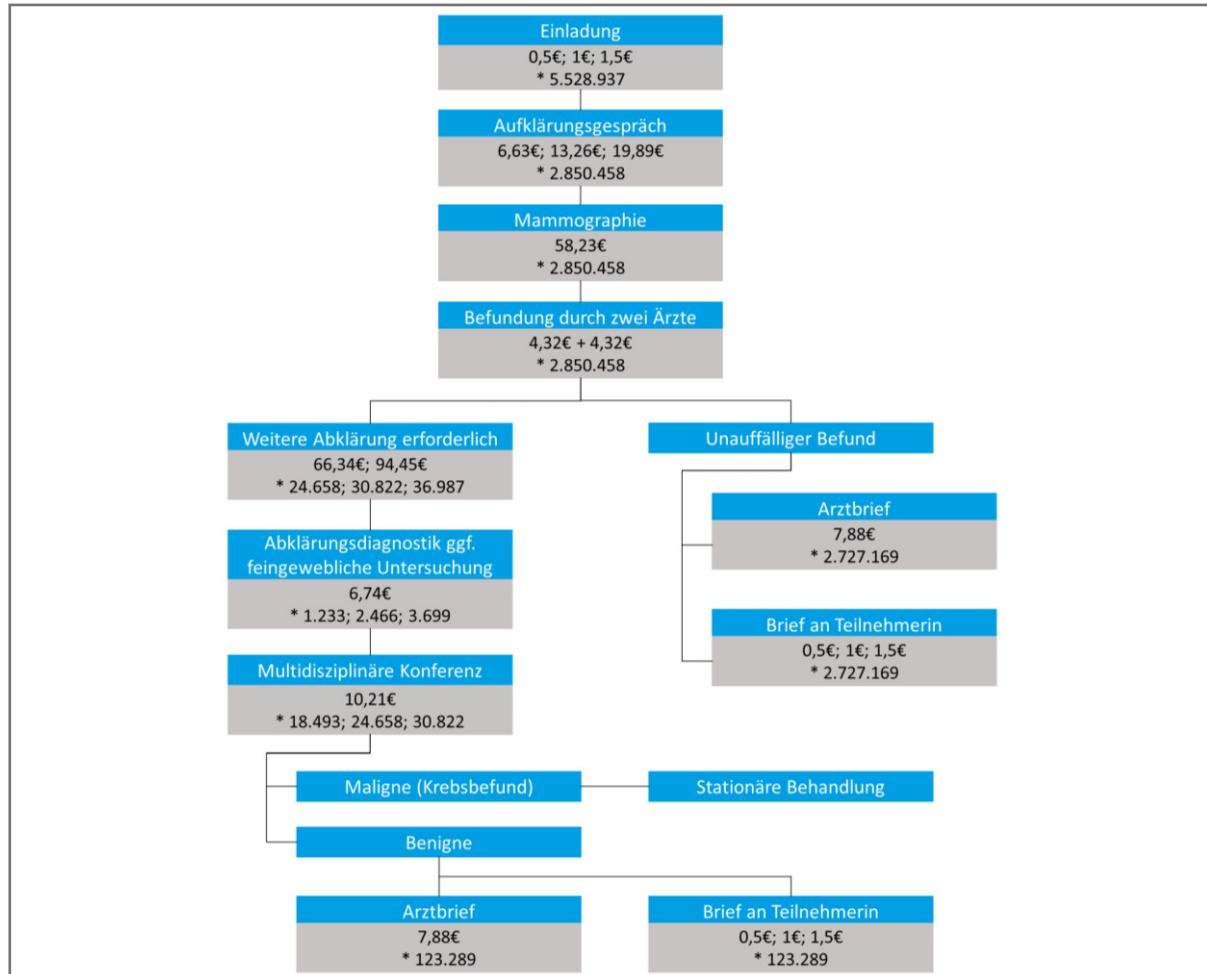
Wie bei jeder Schätzung sind die Annahmen grundlegend für die Validität. Bei dieser Berechnung wurde angenommen, dass alle Teilnehmerinnen gesetzlich versichert sind. Dies ist vertretbar, da der Preis für eine Mammographie laut GOÄ pro Seite und mit 2,3-fachem Satz ca. 40 EUR beträgt. Mit dazugehörigem Aufklärungsgespräch betragen die Kosten für eine beidseitige Mammographie für Privatversicherte mehr als für gesetzlich Versicherte.

Zusätzlich wurde angenommen, dass alle teilnehmenden Ärzte sich an einer wöchentlichen Fallkonferenz beteiligen, in deren Rahmen auffällige Mammographie-Aufnahmen abgeklärt werden. Es wurden drei Szenarien berechnet wobei Szenario 1 eine sparsame, Szenario 2 eine mittlere und Szenario 3 eine teure Schätzung darstellt. Die Szenarien unterscheiden sich bezüglich der Kosten für die Einladung, der Dauer des Aufklärungsgespräches, der Anzahl der Patientinnen, die nach einem positiven Mammographie-Screening eine Stanzbiopsie erhielten und der Anzahl der Ärzte, die an der Behandlung und den wöchentlichen Fallkonferenzen teilnahmen.

Die Anzahl der Stanzbiopsien wurde mit 15 %, 20 % und 25 % der Wiedereinbestellungen (n=123.289) geschätzt. Die Anzahl der teilnehmenden Ärzte wurde auf Grundlage der Ärztestatistik der Kassenärztlichen Bundesvereinigung geschätzt, in dem angenommen wurde, dass für Szenario 1 5 % der Gynäkologen und Radiologen (n=1233) an der Durchführung des Mammographie-Screenings teilnahmen. Szenario 2 rechnete mit 10 % (n=2466) und Szenario 3 mit 15 % (n=3699). Das mittlere Szenario wird mit 268 Mio. EUR als realistischste Schätzung angesehen. Die Anzahl der Stanzbiopsien des Szenarios stimmt ungefähr mit der Anzahl der dokumentierten 15.625 Fälle plus einer Überbehandlung von 3 pro 1000 Frauen überein.

Die Summe für die Durchführung der Mammographie-Aufnahme stellt mit Abstand den größten Kostenblock dar und beträgt unabhängig von den Szenarien 165.982.169 EUR. Aufgrund dieser Berechnungen betragen die Kosten des Mammographie-Programms in Deutschland zwischen 240 und 295 Mio. EUR.

Nr.	Beschreibung	Wert in €	Abrechnungseinheit
1750	Röntgenuntersuchung im Rahmen des Mammographie-Screenings	58,23	Screening beider Brüste
1751	Aufklärungsgespräch im Rahmen des Mammographie-Screenings	6,63	Abrechnung je 5 Minuten, max. 15 Minuten.
1752	Beurteilung von Mammographie-Aufnahmen im Rahmen des Mammographie-Screenings	4,32	Je Patientin
1753	Abklärungsdiagnostik I im Rahmen des Mammographie-Screenings	94,45	Je Patientin
1754	Abklärungsdiagnostik II einschl. ultraschallgestützter Biopsie im Rahmen des Mammographie-Screenings	66,34	Je Patientin
1755	Stanzbiopsie unter Röntgenkontrolle im Rahmen des Mammographie-Screenings	118,36	Je Seite
1756	Histopathologische Untersuchung im Rahmen des Mammographie-Screenings	10,21	Je 3 Stanzen
1758	Teilnahme an einer multidisziplinären Fallkonferenz im Rahmen des Mammographie-Screenings	6,74	Teilnahme persönlich oder telefonisch
1601	Arztbrief	7,88	Je Brief
-	Einladung zum Screening	0,5	Je Brief
-	Brief an Teilnehmerin	0,5	Je Brief

**Tabelle 3:** EBM-Ziffern für das Mammografie-Programm<sup>6</sup>**Abbildung 2:** Zuordnung der EBM-Kennziffern zu den Prozessabläufen eines Mammographie-Screenings<sup>7</sup><sup>6</sup> Quelle: Nach KV Schleswig-Holstein (o. J.).<sup>7</sup> Quelle: Nach KV Schleswig-Holstein (o. J., S. 24).

Szenario 1			
EBM-Nr.	Anzahl	Pauschale in €	Kosten in €
Einladung	5.528.937	0,5	2.764.469
01751	2.850.458	6,63	18.898.537
01750	2.850.458	58,23	165.982.169
01752	5.700.916	4,32	24.627.957
Bei positivem Screening-Ergebnis			
01758	64.124	6,74	432.142
01754	123.289 * 0,85 = 104.796	66,34	6.952.143
01753	123.289 * 0,15 = 18.493	94,45	1.746.697
01756	18.493 / 3 = 6.164	10,21	62.939
Brief an Patientin	2.850.458	0,5	1.425.229
01601	2.850.458	7,88	22.461.609
<b>Summe</b>			<b>245.032.423</b>
Szenario 2			
BM-Nr.	Anzahl Pauschale in €	Kosten in €	
Einladung	Pauschale * 2 = 1€	5.528.937	
01751	Pauschale * 2 = 13,26	37.797.073	
01750	-	165.982.169	
01752	-	24.627.957	
Bei positivem Screening-Ergebnis			
01758	Anzahl * 2 = 128.248	864.389	
01754	123.289 * 0,8 = 98.631	6.543.194	
01753	123.289 * 0,2 = 24.658	2.328.929	
01756	24.658 / 3 = 8.219	83.919	
Brief an Patientin	Pauschale * 2	2.850.458	
01601	-	22.461.609	
<b>Summe</b>		<b>269.068.634</b>	
Szenario 3			
EBM-Nr.	Anzahl Pauschale in €	Kosten in €	
Einladung	Pauschale * 3 = 1,5€	8.293.406	
01751	Pauschale * 3 = 19,89	56.695.610	
01750	-	165.982.169	
01752	-	24.627.957	
Bei positivem Screening-Ergebnis			
01758	Anzahl * 3 = 192.371	1.296.583	
01754	123.289 * 0,75 = 92.467	6.134.244	
01753	123.289 * 0,15 = 30.822	2.911.162	
01756	30.822 / 3 = 10.274	104.898	
Brief an Patientin	Pauschale * 3	4.275.687	
01601	-	22.461.609	
<b>Summe</b>		<b>294.147.850</b>	

Tabelle 4: Berechnung der Kosten für das Mammographie-Screening-Programms<sup>8</sup>

<sup>8</sup> Quelle: Eigene Berechnungen.



## Vergangene Ausgaben

1/2019, Heft 1

**Häckl, D., Kossack, N., Schindler, C., Weinhold, I. & Wende, D.** Weiterentwicklung der Morbiditätsparameter im Morbi-RSA – 7 Thesen und Vorschläge. Diskussionspapier.

**Winkler, J., Schwarz, R., Gantner, T., Nack, D. & Schwarz, M.** Blockchain: Die Demokratisierung des Gesundheitswesens? White Paper zur Funktionsweise und den Erfolgsfaktoren für eine Anwendung der Blockchain im Gesundheitswesen.

2/2019, Heft 2

**Schwarz, M.** Kritische Bewertung des Versorgungsprozesses von MRSA-Patienten. Bearbeitung der Forschungsfragestellung im Rahmen einer Master-Thesis am Ludwig Fresenius Center for Health Care Management and Regulation der HHL Leipzig Graduate School of Management.

## **Impressum**

**Forschungsberichte des Wissenschaftlichen Instituts für Gesundheitsökonomie und Gesundheitssystemforschung**

ISSN 2628-4103

*Herausgeber*

**WIG2 GmbH**

Wissenschaftliches Institut für Gesundheitsökonomie und Gesundheitssystemforschung

Vertreten durch:

Dr. Dennis Häckl, Geschäftsführer

Dr. Thomas Höpfner, Geschäftsführer

Markt 8

04109 Leipzig

Web: <https://www.wig2.de/>

Tel.: + 49 341 39 29 40-0

E-Mail: [forschungsberichte@wig2.de](mailto:forschungsberichte@wig2.de)

*Redaktion*

**Dr. Carsta Militzer-Horstmann**

Leiterin wissenschaftliche Entwicklung & Wissenschaftsmanagement

**Franziska Stutzer**

Studentische Mitarbeiterin Abteilung wissenschaftliche Entwicklung & Wissenschaftsmanagement

*Ausgabe*

3/2019, Heft 3

**Tobias Schäffer** „Statistical illiteracy“ im Gesundheitswesen am Beispiel des Mammographie-Screenings. Seminararbeit an der Martin-Luther-Universität Halle-Wittenberg im Studiengang Wirtschaftswissenschaften

**Hannah Lintener** Die Potenziale und Barrieren der assistierten Heimdialyse im Setting der stationären Altenpflege in Deutschland. Bachelorarbeit an der Hochschule Fulda im Fachbereich Pflege und Gesundheit, Studiengang Gesundheitsökonomie und -politik